



OPEN ACCESS

Using clinical simulation to study how to improve quality and safety in healthcare

Guillaume Lamé, Mary Dixon-Woods

THIS Institute (The Healthcare Improvement Studies Institute), University of Cambridge, Cambridge, UK

Correspondence to
Dr Guillaume Lamé, THIS Institute (The Healthcare Improvement Studies Institute), University of Cambridge, Cambridge CB2 0AH, UK; guillaume.lame@thisinstitute.cam.ac.uk

Received 22 June 2018
Revised 3 September 2018
Accepted 6 September 2018

ABSTRACT

Simulation can offer researchers access to events that can otherwise not be directly observed, and in a safe and controlled environment. How to use simulation for the study of how to improve the quality and safety of healthcare remains underexplored, however. We offer an overview of simulation-based research (SBR) in this context. Building on theory and examples, we show how SBR can be deployed and which study designs it may support. We discuss the challenges of simulation for healthcare improvement research and how they can be tackled. We conclude that using simulation in the study of healthcare improvement is a promising approach that could usefully complement established research methods.

improvement research, using the expertise of the authors.¹²

SIMULATION-BASED RESEARCH DESIGNS

Different research designs are made possible using simulation-based techniques.^{13 14} We start by offering a broad overview of how simulation might be used in research, and then provide a short description and examples of three types of studies (descriptive, theory-testing and generation, and evaluating interventions) that might deploy simulation. We discuss the particular issues that may apply in multicentre studies. Finally, we discuss the potential for combining SBR with other methods in mixed-methods studies.

INTRODUCTION

Simulation in healthcare can be broadly defined as a ‘tool, device, and/or environment (that) mimics an aspect of clinical care’.¹ It has a long history in healthcare education,¹ valued for its ability to reproduce some of the conditions of clinical practice and enable learners to practice in a safe environment. More recently, it has been used as a technique for supporting improvement in healthcare systems and processes, for example, by helping to diagnose problems or test new approaches before they are deployed for real.^{2 3} Given this history, most research is *about* simulation (focusing, eg, on its effectiveness in achieving training or practice goals) rather than *through* simulation.⁴ Despite some recent encouraging examples,^{5 6} the potential of simulation for *conducting* research has remained underexploited. In this article, we identify simulation-based research (SBR) as a distinct research strategy that seeks to generate scientific knowledge about human and organisational behaviour through use of simulation techniques that may take diverse forms, and we explore the role of SBR in the study of how to improve quality and safety in healthcare^{7–11} by offering an informal overview of relevant literature.

We build on the methodological literature in healthcare improvement research to frame our discussion of research designs and strategies, and we draw from the literature on simulation-based training in clinical settings to present the practical side of designing and delivering simulations. As our focus is on simulation for the conduct of research, we do not consider the now large literature on simulation as a training/education method or as a method of undertaking quality improvement. Articles included in the review were selected on the perceived insights they could generate for

Simulation as a way of studying clinical settings

Simulation for research classically seeks to reproduce features of a real-world phenomenon so that it can be studied. In researching how to improve quality and safety in healthcare, an important strategy involves the use of structured scenarios that set up specific settings or events that evoke or replicate features of real-world clinical situations,^{1 15} with the aim of producing data that can be analysed for purposes of generating applied or theoretical knowledge, or both.⁹

A simulator is the *medium* through which a simulation scenario is delivered to participants.^{16 17} Simulators are of many different kinds, including (but not limited to) manikins (which may be more or less interactive), study cases and actors playing the role of patients (so-called ‘simulated patients’).¹⁸ The choice of simulator depends on the research goals and on the clinical situation under study. Generally, the aim is that participants’ reactions and behaviours be as close as possible to what they would experience and do in a real situation, so their perceptions of events, timing, environment and cues are all important.^{16 19}

Simulations may take place either in dedicated simulation centres or in situ, in real-life clinical settings. Recommendations for the design of simulation centres suggest taking a range of factors into consideration, including layout of the different spaces, the design of the control room or noise insulation.²⁰ In situ simulations, which take place in people’s own workplaces, have a number of strengths; for example, one study was able to identify latent safety threats in an emergency department using unannounced simulations that were organised during normal shifts.²¹ However, in situ simulation is not free of challenges: it may, for example, interfere with care or disturb patients and staff.^{21 22}



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY. Published by BMJ.

To cite: Lamé G, Dixon-Woods M. *BMJ Stel* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjstel-2018-000370

Box 1 Randomised trial on the impact of training on speaking-up behaviour

- ▶ Although individuals speaking up about concerns is often the final barrier to adverse events in most high-risk industries, previous research has shown that many hurdles prevent individuals from giving voice healthcare. Raemer *et al*⁸⁸ investigated the impact of a 50 min training workshop on the ability of anaesthesiologists to speak up in the operating room. The intervention group had the 'speaking-up' training before the simulation; for the control group, it was after the simulation. The scenario included three events where anaesthesiologists had the opportunity to speak up to a surgeon, a nurse and another anaesthesiologist, respectively. Simulations were video-recorded, and the reactions of participants to the three events were coded based on the videos. The debriefing sessions were also videotaped, and the hurdles and enablers to speaking up mentioned by participants and passive observers of the simulations were coded based on an existing taxonomy.
- ▶ There was no significant difference between the control group and the experimental group on any of the reactions to the three events. The main hurdle to speaking up was 'uncertainty about the issue'—a surprise for the authors because participants had on average 15 years of experience. In conclusion, the authors acknowledge that the intervention was not effective, and recommend more education on, and institutional support for, speaking up. More generally, these results challenge the efficacy of relying solely on educational interventions to improve speaking-up behaviours.

Various types of data can be collected during a simulation that may be used for research purposes. Simulators themselves often collect data: high-fidelity manikins, for example, can record data such as compression or body position, while laparoscopic simulators can use movement-tracking to assess dexterity.²³ Physiological measurements of participants' reactions may include monitoring of skin resistance as an indicator of the activity of the sympathetic nervous system²⁴ or assessing the cortisol level of participants to measure stress.²⁵ Researchers may also use observational checklists to code the behaviour of simulation participants (eg, box 1).²⁶ Pre-post questionnaires are widely deployed: in a study of the impact of stress on the performance of paramedics, for example, LeBlanc *et al* used questionnaires to measure anxiety before and after simulation.²⁵

Qualitative data from simulations can also be collected either in real time or through use of video recordings. The think-aloud technique, whereby participants verbalise their thoughts during the simulation, is an example of one commonly used method.²⁷ Post-simulation debriefing sessions are also valuable sources of data, as illustrated in box 1. Taking place after a simulation has ended, debriefs may involve feedback to the participants, who are in turn invited to discuss their experiences.^{28 29} One-to-one interviews or focus groups may also be used.⁸

Descriptive studies

One important role for SBR is in understanding what happens in healthcare organisations and why, perhaps by reproducing a situation or condition of interest in a simulation. Researchers may design a simulation to reproduce current practice in a controlled environment in order to observe specific aspects. The aim is to document individual and collective behaviour, and to identify

Box 2 Exploratory study on information sharing in surgical teams

- ▶ Collaboration and teamwork are essential elements of safe care. Cumin *et al*⁸⁹ investigated the sharing of information between individuals in a surgery team. Twenty teams of six people participated in surgical simulation. Before each simulation, participants received an individual case briefing note. Each team member had in their briefing note a specific piece of information that the others did not have, but which was clinically useful for managing the simulated case. Researchers counted if, and when, the information items were shared with the rest of the team. They also tested team members about the information items after the simulation, using a questionnaire.
- ▶ Team members were 5.0 times (95% CI 1.5 to 18.4, $p=0.01$) more likely to remember an information after the simulation if it had been shared during a formal communication time rather than at another time. However, in a significant number of instances information items were not shared (38%), and not all members participated equally (anaesthesiologists and senior surgeons were more likely to share information). The results support the importance of formal communication slots before surgery, but still highlight insufficient information sharing and imbalance between team members.

patterns of interactions and thought processes. For instance, Fossum *et al*³⁰ used written patient cases as a simulator to investigate nurses' thinking strategies when faced with malnutrition and pressure ulcers in nursing homes. Box 2 describes a study on information sharing in surgical teams.

The output can be a *descriptive account*, as, for instance, in a study of how professionals share information in the operating room (box 2), or how nurses make decisions about deteriorating patients.³¹ Used in this way, a distinctive strength of SBR is that it can generate data that could not be obtained with other methods, enabling study of issues that are not possible using other methods for ethical, practical or safety reasons.⁷ It may be especially useful in the study of rare events, where it is not feasible for researchers to wait for a rare event to occur and hope to be there to observe it—and observe it only once. In a simulated environment, rare but potentially catastrophic events or conditions can be reproduced as often as desired, potentially under varying conditions. For instance, simulation has been used to observe how care teams managed malignant hyperthermia, a rare life-threatening condition.³²

Theory-testing and generation

Simulation-based studies can be useful when the aim is to generate, evaluate and extend theories relating to quality and safety. For example, one important hypothesis based on studies in cognition and human biology is that sleep deprivation has adverse effects on the performance of physicians.³³ SBR has made a useful contribution to investigating this hypothesis (box 3). The impact of noise on anaesthetists' stress level in operating theatres has also been studied using simulation.³⁴ Similarly, social and cognitive psychology studies have generated hypotheses about the influences of peer pressure, including the possibility that individuals will conform to a group's opinion even if they think it is false. Simulated experiments with healthcare students have shown how this could affect clinical practice.^{35 36}

Box 3 Randomised trial on the impact of sleep deprivation on non-technical skills

- The impact of sleep deprivation on the performance of anaesthetists is still uncertain, and research so far has mainly focused on technical skills. Neuschwander *et al.*⁷⁹ studied the impact of sleep deprivation on non-technical skills, including 'team working, situation awareness, decision-making, and task management'. The authors developed a crisis management simulation scenario, using a high-fidelity manikin. Also, 10 participants undertook the scenario after a night shift and 10 after a rested night. Two blinded assessors rated the performance of participants using a validated scoring tool.
- The non-technical skills score was significantly lower for the sleep-deprived anaesthesiologists. In particular, team working scores were significantly lower. Self-confidence in anaesthesia skills just before the simulation was also significantly lower in the sleep-deprived group. These findings are important since non-technical skills are suspected to play a key role in avoiding serious adverse events. This study also illustrates the difficulty of recruiting when participation is voluntary: 100 participants were screened, but only 21 agreed to participate. However, the authors argue that the significant difference in non-technical skills makes lack of power unlikely.

Box 4 Uncontrolled before–after study of a new drug packaging system

- Medication errors are a leading cause of adverse events in hospitals. Garcia *et al.*⁹² studied the impact of a new labelling system using a simulated medicine room. For 30 min, each participant was handed a new medication chart once he/she had completed preparation for the previous one. Researchers timed the preparation of each medication chart using a stopwatch and counted the number of errors in preparation with the standard labelling system. They repeated the experience 3 months later, using a new labelling system proposed in the literature by Endestad *et al.*⁹⁹
- The error rate remained low with no significant change, but nurses were significantly quicker in their preparation with the new labelling system. These results contrast with a previous on-screen experiment, where the error rate decreased with the new system.⁹⁹

deployments of simulation in the different study types identified by Portela *et al.*¹³ Some SBR studies use pre-post designs, usually involving an initial observation of the phenomenon of interest performed in a simulated environment, followed by the introduction of an intervention (training, new procedure, new equipment, etc) and repeated measures taken in the same simulated environment (c.f. [box 4](#)).

The ability of SBR to reproduce situations identically before and after increases confidence that the intervention can explain the variation in outcomes of interest. However, in general, simple uncontrolled pre-post studies suffer from weaknesses in attributing causality to the intervention.^{14 37} Time-series designs,

Evaluating interventions

A particularly attractive role for simulation is in evaluating interventions that seek to improve care, not least by affording experimental and other study designs that would otherwise be difficult or impossible to do in real life. [Table 1](#) identifies possible

Table 1 Roles for simulating in different study types aimed at evaluating improvement interventions

Class of studies	Potential role(s) for simulation	Potential data collection method(s)	Example(s)
Quality improvement projects	<ul style="list-style-type: none"> ► Understand the problem situation ► Assess intervention feasibility ► Evaluate the effect of the intervention ► Optimise design and implementation of the intervention 	<ul style="list-style-type: none"> ► Observational checklists ► Focus groups and interviews ► Simulator-collected measures ► Physiological measures ► Questionnaires 	<ul style="list-style-type: none"> ► Combination of simulation and Failure Modes and Effects Analysis in a prospective risk analysis⁹⁰ ► Identification of latent threats in a new hospital facility⁹¹
Effectiveness studies: <i>Randomised controlled trials (RCTs)</i> <i>Quasi-experimental studies</i> <i>Observational studies</i>	<ul style="list-style-type: none"> ► Evaluate the effectiveness of the intervention ► Pre-test interventions 	<ul style="list-style-type: none"> ► Observational checklists ► Simulator-collected measures ► Physiological measures ► Questionnaires 	<ul style="list-style-type: none"> ► RCT: simulation-based trials of surgical checklists,^{6 46} trial on the impact of sleep deprivation on non-technical skills (box 2),⁷⁹ trial on the impact of training on speaking-up behaviour (box 1)⁸⁸ ► Quasi-experimental studies: uncontrolled study of a new drug packaging system (box 4),⁹² uncontrolled evaluation of a paediatric resuscitation training package,⁹³ non-randomised controlled study of an intervention to improve the management of distractions and interruptions during ward rounds⁹⁴ ► Observational studies: simulation-based longitudinal study of crisis resource management ability⁹⁵
Process evaluations	<ul style="list-style-type: none"> ► Evaluate how the intervention is received by participants 	<ul style="list-style-type: none"> ► Observational checklists ► Focus groups and interviews ► Simulator-collected measures ► Physiological measures ► Questionnaires 	<ul style="list-style-type: none"> ► Use of simulated patients as part of a multimethod process evaluation of an intervention to improve youth-friendly services for sexually transmitted infections⁹⁶
Qualitative studies	<ul style="list-style-type: none"> ► Explore perceptions of the intervention ► Produce descriptions and theoretically informed analysis of scenarios 	<ul style="list-style-type: none"> ► Focus groups and interviews ► (Video-)ethnography 	<ul style="list-style-type: none"> ► Qualitative study of the simulation of an audio-visual telehealth service⁹⁷
Economic evaluations	<ul style="list-style-type: none"> ► Feed data to economic models and projections 	<ul style="list-style-type: none"> ► Observational checklists ► Simulator-collected measures ► Physiological measures ► Questionnaires 	<ul style="list-style-type: none"> ► Use of clinical simulation in the cost-effectiveness evaluation of an electronic health record system⁹⁸

Box 5 Multicentre, cross-sectional observational study of paediatric sepsis management

- Severe sepsis is a major source of morbidity and mortality in paediatric patients. Adherence to associated guidelines has been shown to improve outcomes. Kessler *et al*¹⁰⁰ compared the practice of 47 teams in 24 emergency departments (EDs) on the management of paediatric sepsis. The teams came from both paediatric and general EDs. Simulations were conducted in situ, and the primary outcome was adherence to a guideline measured by a six-component checklist. The investigators' hypothesis was that paediatric EDs would adhere to the guideline more than general EDs.
- Paediatric EDs demonstrated greater adherence to the guideline than general EDs. Nonetheless, in an adjusted regression analysis the only factor associated with greater adherence was composite team experience (mean number of years of experience as a medical professional for each team), suggesting that hospitals should give priority to this factor when trying to improve paediatric sepsis management.

where multiple data points are collected before and after the intervention,¹⁴ are generally preferred, but even better, when the aim is to attribute causality, are controlled designs.¹³

Simulation is especially well-suited to facilitating controlled studies, which expose one group but not the other(s) to the intervention(s) (boxes 1 and 3). Controlled studies using simulation have been used, for example, to compare semiautomated defibrillators and automated external defibrillators for the management of in-hospital sudden cardiac arrest.³⁸ Simulation may also have some role in supporting process evaluations, which look at how the intervention is implemented and received.^{39 40}

Multicentre studies

Cheng *et al* provide a guide for multicentre simulation studies,⁴¹ which offer several advantages. The ability to standardise the simulation across participating sites helps to isolate independent variables and to reduce the risk of bias introduced by variations in local contexts.⁴¹ Because multicentre studies provide larger sample sizes, they can increase the generalisability of findings, and enable comparisons between sites.⁴¹ For instance, one study investigated variability in chest compression during paediatric cardiopulmonary resuscitation (CPR) in nine hospitals.⁴² It found that the quality of CPR varied between hospitals even when a just-in-time training intervention was delivered, thus calling into question the effectiveness of current guideline implementation strategies and training approaches for improving discrepancies in care quality. Box 5 shows an example of a multicentre simulation-based study.

Mixed-method strategies

Simulation can be used both as a stand-alone method or as a complement to other research strategies.¹³ In healthcare improvement research, multimethod approaches are important when studying 'dynamic, complex and interacting systems in which innovations in [Quality Improvement] are implemented'.⁴³ For instance, a problematic situation might be analysed through ethnography, structured observation methods and hazard analysis approaches to identify and characterise the problems and opportunities for intervention; simulation may then be used to evaluate a candidate intervention in a pilot study. If the simulation confirms that the intervention impacted outcomes,

these findings would be generalised qualitatively: the causal relationship found in the simulation would be generalised, but not the quantitative strength measured during the simulation.⁴⁴ Finally, a multicentre controlled study could increase external validity. Process evaluation can be used to assess what happened in practice, fidelity to intervention design and implementation challenges.¹³ This kind of combination of methods provides a robust framework for evaluation, allowing triangulation and mitigating the weaknesses of each independent method.⁴⁵ For instance, the impact of surgical safety checklists on care quality has been explored both through simulation,^{6 46} clinical trials^{47 48} and qualitative research.⁴⁹

CONSIDERATIONS IN SIMULATION-BASED RESEARCH

Simulation as a research method has some distinctive advantages. It also poses a number of challenges: practical, pertaining to validity and fidelity, reporting and ethics.

Practicality

The practical organisation of SBR can be challenging. One barrier is the potential cost.⁷ High-fidelity manikins and simulation facilities come at an often high price. Lower-fidelity simulators are less expensive, but may not offer the same possibilities (physical and functional resemblance to real patients, real-time measurements of the manikin's state). Simulation also requires space and facilities, while setting up and managing the simulator may involve specific expertise.⁵⁰ Once the study has been designed, recruitment can also be an issue, especially without leadership support.⁷ The lack of faculty protected time can be a barrier towards participating in or organising simulations.⁵¹

Methodological challenges: validity

SBR poses specific challenges concerning validity compared with more traditional research approaches. *Internal validity* refers to the ability to show that correlations between observations in a study are causal in nature.⁵² In other disciplines like psychology, economy or management research, the strength of behavioural laboratory experiments is often thought to lie in their internal validity.^{52–55} For SBR to have high internal validity, the design and reporting of studies needs to be handled carefully.^{1 56 57} In particular, scenarios need to be carefully described so that simulations are as standardised as possible.⁷ For instance, in the study of a new drug packaging system described in box 4, it was important that the simulations performed before and after the introduction of the new packaging differed as little as possible, so that changes in performance could be attributed to the new packaging system rather than to other elements. If the presentation of the medication charts had been modified, or if the organisation of the drug storage system had been changed, these changes may have affected performance. This would have made it difficult to conclude on the impact of the labelling system.

To ensure consistency in replications of the same scenario, frameworks used to describe scenarios in simulation-based education may, with some adaptations, be suitable for clinical simulation research. For instance, the TEACH Sim framework covers simulation objectives, audience, patient details, simulator details, scenario script (sequence of events and the expected reaction of the learner), equipment, confederates and actors, and team composition.⁵⁸ When the simulation involves confederates or actors (eg, in the roles of patients or other healthcare professionals), they must be trained in order to reproduce the same behaviour in each simulation, and to respect the scenario script.⁷

To ensure that they unfold as intended, scenarios should ideally be pilot-tested.⁵⁹

External validity (applicability or generalisability)^{60–61} describes ‘whether causal relationships can be generalized to different measures, persons, settings, and times’.⁶² Clearly, the generalisability of research findings generated in a simulated environment to ‘real life’ and patient outcomes is an important question.⁷⁹ It is of particular concern in SBR, because, like laboratory experiments,^{52–55} SBR artificially reduces the environmental complexity surrounding the studied phenomenon, potentially weakening confidence in how far it is possible to generalise from these highly controlled conditions.^{63–64} For instance, although some studies specifically address the issue of interruptions in clinical practice,⁶⁵ other studies will attenuate this dimension to put the emphasis on other aspects, such as clinical complications. However, in everyday practice, both factors interact: complications happen and interruptions occur, sometimes simultaneously and other factors such as social relationships also play a role. It is important to know to what extent the results of the study on interruptions would hold in this much richer context.

One way of evaluating the external validity of lab experiments is to compare them with related field studies.^{53–66} In psychological research, the findings suggest that laboratory experiments are best suited to establishing the existence and direction of effects between variables, not the precise magnitude of the effects measured in the lab.^{53–66} This cautious approach has also been recommended in economics.⁴⁴

An alternative view of external validity is that it should depend on the research aims and, since ‘real life’ is not neutral, objective phenomenon, field observations should always be understood as socially constructed and theoretically informed.^{54–63} Finally, when critiques argue that lab experiments create a strong observer effect (a modification of the behaviour created by the presence of an observer, sometimes called the ‘Hawthorne effect’), promoters of lab experiments respond that field research also involves observers, whose presence influences behaviour.⁶³ Therefore, the difference between lab experiments and field-work is less clear-cut than it is sometimes presented.

Construct validity is ‘the degree to which a score can be interpreted as representing the intended underlying construct’.⁶⁷ Construct validity is not a binary notion; it is built with accumulating evidence to support interpretations of assessment data.⁶⁸ SBR does not seem different from other improvement research approaches regarding construct validity, although Cook highlights the need for validity studies on measurement tools used in simulation (eg, scoring tools or checklists).²⁶

Methodological issues: fidelity

Fidelity broadly refers to the extent to which a simulation reproduces the experience of the real-world situation it aims to replicate. The concept is problematic, often ill-defined and used in a binary way (‘high’ or ‘low’ fidelity),^{16–19} whereas a closer analysis shows that the notion is more complex. Certain aspects of the ‘real world’ may be more important in some studies, allowing more freedom and less resemblance on other dimensions.

Beaubien and Baker describe fidelity on three dimensions: psychological, environment and equipment.⁶⁹ Tun *et al* propose to decompose fidelity into three dimensions: patient fidelity (the extent to which the simulator mimics actual patient behaviour), clinical scenario fidelity and healthcare facilities fidelity. Rather than focusing only on high-technology manikins with elaborate physiological responses, these authors insist that all three dimensions of fidelity are important in creating ‘an accurate

representation of real-world cues and stimuli’.¹⁶ Hamstra *et al* insist on *functional task alignment* as a necessary complement to *physical resemblance*.¹⁹ But even physical resemblance can be multifactorial. For instance, very simple and low-cost surgical simulators can be useful, as in the example of chicken breast used to teach ultrasound-guided vascular access.⁷⁰ Chicken breasts do not even remotely look or feel like human bodies, but their physical properties replicate the effect of ultrasound better than high-fidelity manikins.⁷⁰ This can make chicken breasts a more appropriate simulator when the focus is on the ultrasound dimension: they allow for the right actions to be performed (*functional task alignment*) and adequately replicate the physical parameters of interest (*physical resemblance*).¹⁹

As in the use of simulation for education,^{16–19–69} important considerations for fidelity in SBR include the perceived realism/authenticity of the scenario, the patient/simulator and the simulation environment.^{7–16} Researchers need to ensure that the simulation is sufficiently authentic (in terms of physical resemblance and functional similarity)¹⁹ to enable comparisons with participants’ ‘real-life’ behaviour. They should also pay attention to the *phenomenal* (emotions, beliefs) and *semantical* (meaning, theories, information) content of scenarios, before focusing on the *physical* dimension.⁷¹ All three aspects are important and will affect the behaviour of participants.

Though it is sometimes proposed that high-technology manikins might lead to better learning outcomes than less elaborate simulators,¹⁹ evidence of the relationship between the physical fidelity of the simulator and the learning outcomes in simulation-based education is thin.⁷² The choices largely depend on the situation to be studied and the objectives of the simulation. For instance, when studying surgical dexterity, it may be important to reproduce human anatomy, and high-fidelity manikins, cadavers or virtual reality will be appropriate.^{23–73} Similarly, when training for highly technical skills in obstetric emergencies, ‘high-fidelity’ manikins generate better results than simpler ‘doll-like’ manikins.⁷⁴ However, video vignettes, an apparently less sophisticated medium, proved appropriate when studying general practitioners’ decisions to investigate suspected lung cancer,⁷⁵ because they provided enough relevant information and resemblance with the real task. The choice of the setting for the simulation can also be important, for instance, by choosing to organise simulations *in situ*, that is, where care is routinely performed,²² rather than in a simulation lab.

Researchers should of course be aware that the behaviour of participants in a simulation may not reproduce what they would do ‘in real life’. For instance, the level of psychological safety experienced by participants may affect their engagement with the simulation.^{76–77} The style of debriefing may also affect the outcomes, in particular if judgement is involved.²⁹

Reporting

Recent literature reviews have highlighted the shortcomings of quantitative studies on simulation-based educational interventions, in terms of both reporting (simulation context, outcomes, statistical methods) and statistical analysis.^{1–56} This suggests a need for attention to quality of reporting for SBR. Researchers should be aware of extensions to the Consolidated Standards of Reporting Trials statement for randomised trials and the Strengthening the Reporting of Observational Studies in Epidemiology statement for observational studies.⁷⁸ These guidelines recommend that researchers should specify whether they are reporting *research on simulation*, or use *simulation as an investigative method*. Further, the guidelines suggest that

generalisability of the findings from simulation outcomes to patient outcomes should be explicitly discussed.

Ethics

A particular strength of SBR is that it enables studies of quality and safety without putting real patients (or staff) at risk. For instance, it might not be acceptable to assess the effects of sleep deprivation on a physician's performance directly with real patients, but SBR has been used to evaluate the link between sleep duration and the skills of surgeons or anaesthesiologists (box 3).^{23 79} New procedures and systems can be evaluated in a safe environment without threatening patient care, for example, allowing identification that a new drug labelling system could have generated more medication errors (box 4). Similarly, scenarios can be allowed to unfold, even when a researcher has noticed a dangerous situation developing—something that would be unacceptable in direct observation (box 1).

SBR is not free of ethical consequence, however. For instance, during in situ simulation, the person recording should avoid filming bystanders, as this raises privacy issues.²² Perhaps less straightforward are issues relating to adverse outcomes and the use of deception.^{80–83} Calhoun *et al* report a simulation scenario combining both issues, where a confederate acting as a senior clinician ordered the administration of the wrong medication for the simulated case.⁸¹ The objective was that the participants would challenge this instruction even though it meant overcoming hierarchical distance. If they went on to administer the wrong drug, the simulated patient died. This kind of deception can be problematic because participants are asked to 'suspend disbelief' during simulation, but are then confronted with potentially traumatising outcomes (death and serious injury, eg, or evidence of their own culpability and fallibility).⁸⁰ Though some argue that poor outcomes during simulation may promote learning,⁸² the potential for psychological harm is there, not just for the intended targets of the improvement but also for other simulation stakeholders such as standardised patient actors or individuals managing the simulation.⁸⁴ One argument is that these risks are acceptable if the right precautions are taken (a safe environment, both physically and psychologically, a good debriefing, avoiding death scenarios with early learners, providing follow-up after simulation or even mentioning the possibility of patient death during pre-briefing).^{81–83} Others argue for careful reflection on the objectives of the simulation and participants' profiles.⁸⁰

CONCLUSIONS

This paper provides an overview of how simulation can be used in research aimed at studying how to improve quality and safety in healthcare. We propose that simulation has considerable potential; examples thus far show that it is capable of accommodating multiple research designs and allowing study of some situations that would otherwise escape scientific evaluation, such as rare events or situations where direct observation or experiments are not possible for ethical or other reasons. Simulation has the advantage of providing researchers with a controlled environment to test their hypotheses, and to do so safely for patients and participants. It allows for controlled variation of variables, a foundational element of empiric scientific knowledge.⁴⁵

Given these strengths, it is no surprise that clinical simulation networks are including simulation-based research in their programmes of work beside simulation-based education and training,^{4 5} and that the use of simulation to improve quality and safety in healthcare is also gaining increased attention on

Box 6 Recommendations

- ▶ Simulation is a flexible and pluripotent technique that can be used in multiple study designs in healthcare improvement research.
- ▶ Researchers should consider simulation-based research when it is otherwise difficult, costly or ethically impossible to obtain data about a phenomenon.
- ▶ Researchers should evaluate the possibility of integrating simulation as part of a mixed-methods design, where the data obtained in the controlled environment of the simulation can be triangulated with empirical field data.
- ▶ Simulation-based research raises distinctive issues relating to study design, methods, and ethics that require ongoing attention.

the agenda of the simulation community.^{85 86} In this context, research *through* simulation⁴ opens promising perspectives for healthcare improvement research (box 6).

Despite all its merits, SBR is not without its disadvantages. Its very artificiality affects the conclusions that can be drawn from the results of simulation studies, for example, in limiting what can be learnt about context. Overall, the influence of the simulation environment on the performance of participants is still relatively unknown,⁹ so caution is needed in drawing more general conclusions from simulation-based studies. Several authors highlight the need to match simulation design with simulation objectives,^{16 19 69} which supports the nuanced and reflexive approach to applicability/generalisability expressed in other disciplines.⁶³ Until more is known, it seems sensible to generalise (with caution) the existence and direction of a relationship between variables established during simulation, but not the strength of this relationship.⁴⁴

The review offered here sought to deepen understanding of a complex emerging area of research practice, and did not seek to be systematic. Transparency on the selection of articles⁸⁷ is therefore limited. Future studies could complement our approach with a more systematic review of SBR in healthcare improvement research. This would help to further assess the current state of simulation-based improvement research.

In conclusion, simulation has the potential to become a useful addition to improvement researchers' methodological toolkits. The main value simulation can bring to improvement research is by helping collect data on phenomena that researchers can hardly observe. In this way, simulation can help describe individual and organisational behaviour, generate theory and evaluate improvement interventions.

Contributors MD-W had the original idea for the paper and both authors developed the content. GL surveyed the literature and wrote the first draft of the paper, and MD-W adapted and revised this draft. Both authors approved the submitted version.

Funding GL and MD-W are supported by the Health Foundation's grant to the University of Cambridge for The Healthcare Improvement Studies (THIS) Institute. THIS Institute is supported by the Health Foundation, an independent charity committed to bringing about better health and health care for people in the UK. This work was also supported by MDW's Wellcome Trust Investigator award WT09789. MDW is a National Institute for Health Research (NIHR) Senior Investigator.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any

purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

REFERENCES

- Cook DA, Hatala R, Brydges R, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA* 2011;306:978–88.
- Cheng A, Grant V, Auerbach M. Using simulation to improve patient safety: dawn of a new era. *JAMA Pediatr* 2015;169:419.
- Slakey DP, Simms ER, Rennie KV, et al. Using simulation to improve root cause analysis of adverse surgical outcomes. *Int J Qual Health Care* 2014;26:144–50.
- Dieckmann P, Phero JC, Isenberg SB, et al. The first research consensus summit of the society for simulation in healthcare: conduction and a synthesis of the results. *Simul Healthc* 2011;6:S1–9.
- Cheng A, Auerbach M, Calhoun A, et al. Building a community of practice for researchers: the international network for simulation-based pediatric innovation, research and education. *Simul Healthc* 2018;13:S28–34.
- Pucher PH, Aggarwal R, Qurashi M, et al. Randomized clinical trial of the impact of surgical ward-care checklists on postoperative care in a simulated environment. *Br J Surg* 2014;101:1666–73.
- Cheng A, Auerbach M, Hunt EA, et al. Designing and conducting simulation-based research. *Pediatrics* 2014;133:1091–101.
- Munroe B, Buckley T, Curtis K, et al. Designing and implementing full immersion simulation as a research tool. *Australas Emerg Nurs J* 2016;19:90–105.
- LeBlanc VR, Manser T, Weinger MB, et al. The study of factors affecting human and systems performance in healthcare using simulation. *Simul Healthc* 2011;6:S24–9.
- Marshall M, Pronovost P, Dixon-Woods M. Promotion of improvement as a science. *Lancet* 2013;381:419–21.
- The Health Foundation. *Improvement science. Evidence scan*. 22. London: The Health Foundation, 2011.
- Dixon-Woods M. Systematic reviews and qualitative methods. In: Silverman D, ed. *Qualitative research: theory, method and practice*. 3rd edn. London: Sage, 2010:331–46.
- Portela MC, Pronovost PJ, Woodcock T, et al. How to study improvement interventions: a brief overview of possible study types. *BMJ Qual Saf* 2015;24:325–36.
- Eccles M, Grimshaw J, Campbell M, et al. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Health Care* 2003;12:47–52.
- Gaba DM. The future vision of simulation in healthcare. *Simul Healthc* 2007;2:126–35.
- Tun JK, Alinier G, Tang J, et al. Redefining simulation fidelity for healthcare education. *Simul Gaming* 2015;46:159–74.
- Dieckmann P, Rall M. Simulators in anaesthetic training to enhance patient safety. In: Cashman J, Grounds M, eds. *Recent Advances in Anaesthesia and Intensive Care*. Cambridge: Cambridge University Press, 2007:213–32.
- Chiniara G, Cole G, Brisbin K, et al. Simulation in healthcare: a taxonomy and a conceptual framework for instructional design and media selection. *Med Teach* 2013;35:e1380–95.
- Hamstra SJ, Brydges R, Hatala R, et al. Reconsidering fidelity in simulation-based training. *Acad Med* 2014;89:387–92.
- Kutzin JM. Simulation Design Considerations 2.0: optimizing space and operations. *Clin Simul Nurs* 2016;12:187–96.
- Patterson MD, Geis GL, Falcone RA, et al. In situ simulation: detection of safety threats and teamwork training in a high risk emergency department. *BMJ Qual Saf* 2013;22:468–77.
- Patterson MD, Blike GT, Nadkarni VM, et al. In Situ Simulation: Challenges and Results. In: Henriksen K, Battles JB, Keyes MA, eds. *Advances in patient safety: new directions and alternative approaches. Vol 3: Performance and Tools*. Rockville (MD): Agency for Healthcare Research and Quality (US), 2008.
- Taffinder NJ, McManus IC, Gul Y, et al. Effect of sleep deprivation on surgeons' dexterity on laparoscopy simulator. *Lancet* 1998;352:1191.
- Schuetz M, Gockel I, Beardi J, et al. Three different types of surgeon-specific stress reactions identified by laparoscopic simulation in a virtual scenario. *Surg Endosc* 2008;22:1263–7.
- Leblanc VR, Regehr C, Tavares W, et al. The impact of stress on paramedic performance during simulated critical events. *Prehosp Disaster Med* 2012;27:369–74.
- Cook DA. One drop at a time: research to advance the science of simulation. *Simul Healthc* 2010;5:1–4.
- Thackray D, Roberts L. Exploring the clinical decision-making used by experienced cardiorespiratory physiotherapists: A mixed method qualitative design of simulation, video recording and think aloud techniques. *Nurse Educ Today* 2017;49:96–105.
- Fanning RM, Gaba DM. The role of debriefing in simulation-based learning. *Simul Healthc* 2007;2:115–25.
- Rudolph JW, Simon R, Rivard P, et al. Debriefing with good judgment: combining rigorous feedback with genuine inquiry. *Anesthesiol Clin* 2007;25:361–76.
- Fossum M, Alexander GL, Göransson KE, et al. Registered nurses' thinking strategies on malnutrition and pressure ulcers in nursing homes: a scenario-based think-aloud study. *J Clin Nurs* 2011;20(17–18):2425–35.
- Bucknall TK, Forbes H, Phillips NM, et al. An analysis of nursing students' decision-making in teams during simulations of acute patient deterioration. *J Adv Nurs* 2016;72:2482–94.
- Cain CL, Riess ML, Gettrust L, et al. Malignant hyperthermia crisis: optimizing patient outcomes through simulation and interdisciplinary collaboration. *AORN J* 2014;99:300–11.
- Harrison Y, Horne JA. The impact of sleep deprivation on decision making: a review. *J Exp Psychol Appl* 2000;6:236–49.
- McNeer RR, Bennett CL, Dudaryk R. Intraoperative noise increases perceived task load and fatigue in anesthesiology residents: a simulation-based study. *Anesth Analg* 2016;122:2068–81.
- Kaba A, Beran TN. Impact of peer pressure on accuracy of reporting vital signs: An interprofessional comparison between nursing and medical students. *J Interprof Care* 2016;30:116–22.
- Kaba A, Beran TN, White D. Accuracy of interpreting vital signs in simulation: An empirical study of conformity between medical and nursing students. *J Interprof Educ Pract* 2016;3:9–18.
- Goodacre S. Uncontrolled before-after studies: discouraged by Cochrane and the EMJ. *Emerg Med J* 2015;32:507–8.
- Kobayashi L, Dunbar-Viveiros JA, Sheahan BA, et al. In situ simulation comparing in-hospital first responder sudden cardiac arrest resuscitation using semiautomated defibrillators and automated external defibrillators. *Simul Healthc* 2010;5:82–90.
- Hulscher ME, Laurant MG, Grol RP. Process evaluation on quality improvement interventions. *Qual Saf Health Care* 2003;12:40–6.
- Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: medical research council guidance. *BMJ* 2015;350:h1258.
- Cheng A, Kessler D, Mackinnon R, et al. Conducting multicenter research in healthcare simulation: Lessons learned from the INSPIRE network. *Adv Simul* 2017;2.
- Cheng A, Hunt EA, Grant D, et al. Variability in quality of chest compressions provided during simulated cardiac arrest across nine pediatric institutions. *Resuscitation* 2015;97:13–19.
- Alexander JA, Hearld LR. The science of quality improvement implementation: developing capacity to make a difference. *Med Care* 2011;49:S6–20.
- Kessler JB, Vesterlund L. The External Validity of Laboratory Experiments: The Misleading Emphasis on Quantitative Effects. In: Fréchet GR, Schotter A, eds. *Handbook of Experimental Economic Methodology*. Oxford: Oxford University Press, 2015:391–406.
- Falk A, Heckman JJ. Lab experiments are a major source of knowledge in the social sciences. *Science* 2009;326:535–8.
- Arriaga AF, Bader AM, Wong JM, et al. Simulation-Based trial of surgical-crisis checklists. *N Engl J Med* 2013;368:246–53.
- Reames BN, Krell RW, Campbell DA, et al. A checklist-based intervention to improve surgical outcomes in Michigan: evaluation of the keystone surgery program. *JAMA Surg* 2015;150:208.
- Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med* 2009;360:491–9.
- Bergs J, Lambrechts F, Simons P, et al. Barriers and facilitators related to the implementation of surgical safety checklists: a systematic review of the qualitative evidence. *BMJ Qual Saf* 2015;24:776–86.
- Schmidt E, Goldhaber-Fiebert SN, Ho LA, et al. Simulation exercises as a patient safety strategy: a systematic review. *Ann Intern Med* 2013;158:426.
- Korndorff JR, Arora S, Sevdalis N, et al. The American College of Surgeons/ Association of Program directors in surgery national skills curriculum: adoption rate, challenges and strategies for effective implementation into surgical residency programs. *Surgery* 2013;154:13–20.
- Roe BE, Just DR. Internal and external validity in economics research: tradeoffs between experiments, field experiments, natural experiments, and field data. *Am J Agric Econ* 2009;91:1266–71.
- Mitchell G. Revisiting truth or triviality: the external validity of research in the psychological laboratory. *Perspect Psychol Sci* 2012;7:109–17.
- Schram A. Artificiality: The tension between internal and external validity in economic experiments. *J Eco Method* 2005;12:225–37.
- Scandura TA, Williams EA. Research methodology in management: current practices, trends, and implications for future research. *Acad Manage J* 2000;43:1248–64.
- Zendejas B, Brydges R, Wang AT, et al. Patient outcomes in simulation-based medical education: a systematic review. *J Gen Intern Med* 2013;28:1078–89.
- Pucher PH, Tamblyn R, Boorman D, et al. Simulation research to enhance patient safety and outcomes: recommendations of the innovative patient safety domain group. *BMJ Simul Technol Enhanc Learn* 2017;3(Suppl 1):S3–7.
- Benishke LE, Lazzara EH, Gaught WL, et al. The Template of Events for Applied and Critical Healthcare Simulation (TEACH Sim): a tool for systematic simulation scenario design. *Simul Healthc* 2015;10:21–30.

- 59 Alinier G. Developing high-fidelity health care simulation scenarios: a guide for educators and professionals. *Simul Gaming* 2011;42:9–26.
- 60 Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet* 2005;365:82–93.
- 61 Glasgow RE. What types of evidence are most needed to advance behavioral medicine? *Ann Behav Med* 2008;35:19–25.
- 62 Steckler A, McLeroy KR. The importance of external validity. *Am J Public Health* 2008;98:9–10.
- 63 Visser G, Heyne G, Peters V, et al. The validity of laboratory research in social and behavioral science. *Quality and Quantity* 2001;35:129–45.
- 64 Levitt SD, List JA. Viewpoint: On the generalizability of lab behaviour to the field: Generalizability of lab behaviour to the field. *Can J Eco/Revue canadienne d'économie* 2007;40:347–70.
- 65 Jones J, Wilkins M, Caird J, et al. An experimental study on the impact of clinical interruptions on simulated trainee performances of central venous catheterization. *Adv Simul* 2017;2:5.
- 66 Anderson CA, Lindsay JJ, Bushman BJ. Research in the psychological laboratory: truth or triviality? *Current Dir Psych Sci* 1999;8:3–9.
- 67 Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med* 2006;119:166.e7–16.
- 68 Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ* 2003;37:830–7.
- 69 Beaubien JM, Baker DP. The use of simulation for training teamwork skills in health care: how low can you go? *Qual Saf Health Care* 2004;13:i51–6.
- 70 Fürst RV, Polimanti AC, Galego SJ, et al. Ultrasound-guided vascular access simulator for medical training: proposal of a simple, economic and effective model. *World J Surg* 2017;41:681–6.
- 71 Dieckmann P, Gaba D, Rall M. Deepening the theoretical foundations of patient simulation as social practice. *Simul Healthc* 2007;2:183–93.
- 72 Norman G, Dore K, Grierson L. The minimal relationship between simulation fidelity and transfer of learning. *Med Educ* 2012;46:636–47.
- 73 Grantcharov TP, Bardram L, Funch-Jensen P, et al. Laparoscopic performance after one night on call in a surgical department: prospective study. *BMJ* 2001;323:1222–3.
- 74 Crofts JF, Bartlett C, Ellis D, et al. Training for shoulder dystocia: a trial of simulation using low-fidelity and high-fidelity mannequins. *Obstet Gynecol* 2006;108:1477–85.
- 75 Sheringham J, Sequeira R, Myles J, et al. Variations in GPs' decisions to investigate suspected lung cancer: a factorial experiment using multimedia vignettes. *BMJ Qual Saf* 2017;26:449–59.
- 76 Roussin CJ, Larraz E, Jamieson K, et al. Psychological safety, self-efficacy, and speaking up in interprofessional health care simulation. *Clin Simul Nurs* 2018;17:38–46.
- 77 Rudolph JW, Raemer DB, Simon R. Establishing a safe container for learning in simulation: the role of the presimulation briefing. *Simul Healthc* 2014;9:339–49.
- 78 Cheng A, Kessler D, Mackinnon R, et al. Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE Statements. *Simul Healthc* 2016;11:238–48.
- 79 Neuschwander A, Job A, Younes A, et al. Impact of sleep deprivation on anaesthesia residents' non-technical skills: a pilot simulation-based prospective randomized trial. *Br J Anaesth* 2017;119:125–31.
- 80 Truog RD, Meyer EC. Deception and death in medical simulation. *Simul Healthc* 2013;8:1–3.
- 81 Calhoun AW, Boone MC, Miller KH, et al. Case and commentary: using simulation to address hierarchy issues during medical crises. *Simul Healthc* 2013;8:13–19.
- 82 Gaba DM. Simulations that are challenging to the psyche of participants: how much should we worry and about what? *Simul Healthc* 2013;8:4–7.
- 83 Corveto MA, Taekman JM. To die or not to die? A review of simulated death. *Simul Healthc* 2013;8:8–12.
- 84 Smith AB, Lammers SE. The Ethics of Simulation. In: Palaganas JC, Maxworthy JC, Epps CA, eds. *Defining excellence in simulation programs*. First edn. Philadelphia: Wolters Kluwer, 2015:592–6.
- 85 Scerbo MW, Calhoun AW, Paige JT, et al. The Second society for simulation in healthcare research summit: beyond our boundaries. *Simul Healthc* 2018;13:S1–6.
- 86 Paige JT, Terry Fairbanks RJ, Gaba DM. Priorities related to improving healthcare safety through simulation. *Simul Healthc* 2018;13:S41–50.
- 87 Greenhalgh T, Thorne S, Malterud K. Time to challenge the spurious hierarchy of systematic over narrative reviews? *Eur J Clin Invest* 2018;48:e12931.
- 88 Raemer DB, Kolbe M, Minehart RD, et al. Improving Anesthesiologists' ability to speak up in the operating room: a randomized controlled experiment of a simulation-based intervention and a qualitative analysis of hurdles and enablers. *Acad Med* 2016;91:530–9.
- 89 Cumin D, Skilton C, Weller J. Information transfer in multidisciplinary operating room teams: a simulation-based observational study. *BMJ Qual Saf* 2017;26:209–16.
- 90 Nielsen DS, Dieckmann P, Mohr M, et al. Augmenting Health Care Failure Modes and Effects Analysis With Simulation. *Simul Healthc* 2014;9:48–55.
- 91 Medwid K, Smith S, Gang M. Use of in-situ simulation to investigate latent safety threats prior to opening a new emergency department. *Saf Sci* 2015;77:19–24.
- 92 Garcia BH, Elenjor R, Bjornstad C, et al. Safety and efficiency of a new generic package labelling: a before and after study in a simulated setting. *BMJ Qual Saf* 2017;26:817–23.
- 93 Ersdal HL, Vossius C, Bayo E, et al. A one-day helping babies breathe course improves simulated performance but not clinical management of neonates. *Resuscitation* 2013;84:1422–7.
- 94 Thomas I, Nicol L, Regan L, et al. Driven to distraction: a prospective controlled study of a simulated ward round experience to improve patient safety teaching for medical students. *BMJ Qual Saf* 2015;24:154–61.
- 95 Clarke S, Horeczko T, Carlisle M, et al. Emergency medicine resident crisis resource management ability: a simulation-based longitudinal study. *Med Educ Online* 2014;19:25771.
- 96 Renju J, Andrew B, Nyalali K, et al. A process evaluation of the scale up of a youth-friendly health services initiative in northern Tanzania. *J Int AIDS Soc* 2010;13:32.
- 97 Clay-Williams R, Baysari M, Taylor N, et al. Service provider perceptions of transitioning from audio to video capability in a telehealth system: a qualitative evaluation. *BMC Health Serv Res* 2017;17:558.
- 98 Ben-Assuli O, Ziv A, Sagi D, et al. Cost-Effectiveness Evaluation of EHR: simulation of an abdominal aortic aneurysm in the emergency department. *J Med Syst* 2016;40:141.
- 99 Endestad T, Wortinger LA, Madsen S, et al. Package design affects accuracy recognition for medications. *Hum Factors* 2016;58:1206–16.
- 100 Kessler DO, Walsh B, Whitfill T, et al. Disparities in adherence to pediatric sepsis guidelines across a spectrum of emergency departments: a multicenter, cross-sectional observational in situ simulation study. *J Emerg Med* 2016;50:403–15.